The Ruth H. Hooker Research Library

and Technical Information Center

Building Electronic Journal Collections: Implementing Elsevier Electronic Subscriptions

Laurie E. Stackpole and Roderick D. Atkinson

U.S. Naval Research Laboratory

Project Briefing Presented at the Coalition for Networked Information Fall Meeting, December 1996.

Introduction

To deliver electronic documents to the desktop, the Ruth H. Hooker Research Library and Technical Information Center has developed TORPEDO (The Optical Retrieval Project: Electronic Documents Online). TORPEDO uses commercial software, the EFS (Electronic Filing System) WebFile from Excalibur Technologies. The EFS software provides end users with the ability to search the full-text of documents, using a fuzzy search algorithm. It also provides the ability to browse document collections, say all the articles in a journal issue, and to go directly to a particular article if the volume number and page are known. Document images can be viewed online or printed locally. TORPEDO has been available to NRL users nationally since September 1995.

InfoWeb and TORPEDO

TORPEDO is available through the Library's World Wide Web Pages known as InfoWeb [http://infoweb.nrl.navy.mil]. InfoWeb serves not only as a point of entry for TORPEDO, but also for many other Library services. As an interface to TORPEDO, InfoWeb provides a functional front end that tells the user about the system, provides context-sensitive online help, and launches the Web-based TORPEDO interface, using a customized interface to EFS WebFile. Besides a Web browser, the only software required on the users workstation is a group IV TIFF image viewer. Hyperlinks to suitable shareware and freeware image viewers for PCS, Macs, and Unix systems, as well as browser configuration instructions, are provided on the TORPEDO Web page.

Digital Report and Journal Collections

TORPEDO provides NRL researchers with desktop access to scientific publications previously only available within the Library. Technical reports published by DoD, and DoD contractors in industry and academia, are one important source of information for the NRL user community. Since 1988, the NRL Library has been digitizing and storing optically its large collection of technical reports. This digital collection now consists of 165,000 technical reports (about 9 million pages stored as TIFF images). Reports without security restrictions or limitations on access are being OCRed (optical character recognized) and added to TORPEDO as a key component of the Library's digital collections.

Journal articles are the other information source with which the Library is principally concerned. In 1993, the NRL Library and the American Physical Society agreed to work together to test the electronic dissemination of current journals, *Physical Review Letters* and *Physical Review E*. While the availability of two journals has enabled the Library to test TORPEDO and demonstrate its capabilities, it did not provide the critical mass of information needed to move researchers beyond casual interest. To provide that critical mass, NRL become one of the pilot libraries, the first in the federal sector, to participate in the Elsevier Electronic Subscription Project. The 160 Elsevier journals are currently licensed for this project have been available through TORPEDO since May.

Elsevier Electronic Subscriptions

Elsevier Electronic Subscriptions consist of TIFF images, unedited ASCII text, and bibliographic data in tagged SGML format. The NRL Library has elected to receive and retain this information on CD-ROMs. The Elsevier journals are shipped to NRL on 6 CD-ROMs per month. A short program has been written to mount the Elsevier journals in native CD-ROM format, point to the defined directory structure, and index the full-text so that it is concurrently searchable along with the other documents in TORPEDO. Once the CD-ROMs are received by the Library, the new journals are available to the NRL/ONR community at 2 am the next day, after the index has been rebuilt and restarted.

The software chosen to mount and manage the Library's 500 platter Panasonic CDROM jukebox is from Tracer Technologies. The usual problem with jukebox technology is that multiuser access is slow. Firsttime access to an article requires that the appropriate CD be mounted, which, because this is a mechanical process, can take as much as 18 seconds for a single user. To counter this slow response time, a large 100GB RAID level 5 system was put in place to cache the most frequently requested documents. The Tracer software complies with Hierarchical Storage Management level specification, so that data are migrated from the slower optical media to hard disks where they are left until no additional storage space if available. This level of file migration can increase system response time by an order of magnitude. Once the cache is full, the documents that have been accessed least recently are automatically deleted from the RAID. This arrangement allows the Library to use low cost, but also slower, near-line storage in the from of optical jukeboxes, while assuring rapid access (>10 ms) to frequently requested documents.

Portable Document Files vs. TIFF

Although group IV TIFF continues to be the standard format for imaging systems, the Web community is rapidly embracing PDF (Portable Document Files) as a de facto standard for document dissemination. Many scientific publishers have also moved to PDF as the primary method for disseminating their publications in electronic format. The Library now sees PDF files as both easier to receive and to disseminate and plans in the very near future to discontinue the dissemination of TIFF images and implement Adobe PDF

End-user Perspective

While many publishers are making journal information available through their own Web sites, each journal or the Web-published journals of each publisher, must be searched independently. TORPEDO, on the other hand, allows a researcher to search across all the journals in the Library's collection that are available electronically. While access to many journals to which the NRL Library subscribes is currently available through InfoWeb hyperlinks, this is considered an interim measure that only partially meets the needs of the user community. In addition to requiring that a researcher know the publisher of a particular journal, it assumes that the researcher has narrowed the search to the point of knowing where to look for the information. Configuration issues, such as passwords and IP addresses, have also been difficult to negotiate with vendors for implementation on InfoWeb. In addition, response time from publisher servers has been found to be inadequate. Negotiations are therefore underway with a number of publishers providing site access to their journals on the Web to add these journals to TORPEDO instead.

Conclusion

The NRL Library is well on the way to providing its research community with access to its digital collections through TORPEDO. TORPEDO offers powerful search and retrieval tools that enable end users seated at their computers or workstations to use Web browsers to search, view, and print the contents of large collections of library materials.

| Home | NRL | Suggest | Search |

webmaster@library.nrl.navy.mil
Updated: 13-Mar-97
Maintained by: Erik M. Schwartz